



# wwPDB X-ray Structure Validation Summary Report

Dec 13, 2022 – 12:10 pm GMT

PDB ID : 7ZIM  
Title : JC Polyomavirus VP1 in complex with 3'-Sialyllactose glycomacromolecules  
(aromatic linker)  
Authors : Freytag, J.; Mueller, J.C.; Stehle, T.  
Deposited on : 2022-04-08  
Resolution : 1.55 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the  symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references](#) ) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : **FAILED**  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.31.3

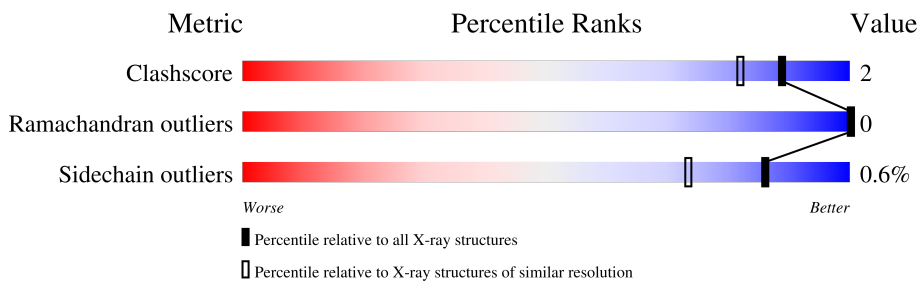
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.55 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	1529 (1.56-1.56)
Ramachandran outliers	138981	1498 (1.56-1.56)
Sidechain outliers	138945	1495 (1.56-1.56)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ .

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	AAA	272	93% 5%
1	BBB	272	90% 6%
1	CCC	272	92% 5%
1	DDD	272	91% 6%
1	EEE	272	91% 5%
2	BaB	3	100%
2	CaC	3	33% 67%
3	A	2	50% 50%

## 2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 12254 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Major capsid protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	AAA	259	2040	1285	349	395	11	0	7	0
1	BBB	256	2045	1286	352	396	11	0	10	0
1	CCC	265	2169	1358	369	430	12	0	16	0
1	DDD	265	2137	1340	368	417	12	0	13	0
1	EEE	259	2050	1291	353	395	11	0	5	0

There are 20 discrepancies between the modelled and reference sequences:

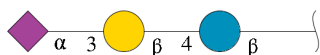
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	18	GLY	-	expression tag	UNP P03089
AAA	19	SER	-	expression tag	UNP P03089
AAA	20	HIS	-	expression tag	UNP P03089
AAA	21	MET	-	expression tag	UNP P03089
BBB	18	GLY	-	expression tag	UNP P03089
BBB	19	SER	-	expression tag	UNP P03089
BBB	20	HIS	-	expression tag	UNP P03089
BBB	21	MET	-	expression tag	UNP P03089
CCC	18	GLY	-	expression tag	UNP P03089
CCC	19	SER	-	expression tag	UNP P03089
CCC	20	HIS	-	expression tag	UNP P03089
CCC	21	MET	-	expression tag	UNP P03089
DDD	18	GLY	-	expression tag	UNP P03089
DDD	19	SER	-	expression tag	UNP P03089
DDD	20	HIS	-	expression tag	UNP P03089
DDD	21	MET	-	expression tag	UNP P03089
EEE	18	GLY	-	expression tag	UNP P03089
EEE	19	SER	-	expression tag	UNP P03089
EEE	20	HIS	-	expression tag	UNP P03089

*Continued on next page...*

Continued from previous page...

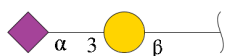
Chain	Residue	Modelled	Actual	Comment	Reference
EEE	21	MET	-	expression tag	UNP P03089

- Molecule 2 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-4)-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	BaB	3	43	23	1	19	0	0	0
2	CaC	3	32	17	1	14	0	0	1

- Molecule 3 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	A	2	21	11	1	9	0	0	1

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C<sub>4</sub>H<sub>10</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	1	Total C O 7 4 3	0	0
4	EEE	1	Total C O 7 4 3	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	AAA	308	Total O 308 308	0	0
5	BBB	320	Total O 320 320	0	0
5	CCC	365	Total O 365 365	0	0
5	DDD	383	Total O 383 383	0	0
5	EEE	327	Total O 327 327	0	0

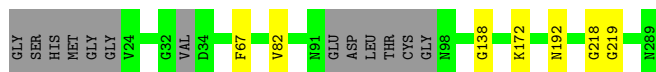
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

Note EDS failed to run properly.

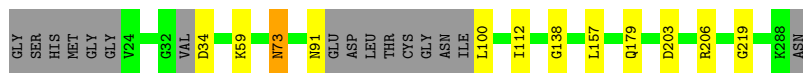
- Molecule 1: Major capsid protein VP1

Chain AAA:  93% • 5%



- Molecule 1: Major capsid protein VP1

Chain BBB:  90% • 6%



- Molecule 1: Major capsid protein VP1

Chain CCC:  92% • 5%



- Molecule 1: Major capsid protein VP1

Chain DDD:  91% • 6%




- Molecule 1: Major capsid protein VP1

Chain EEE:  91% • 5%



- Molecule 2: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-4)-beta-D-glucopyranose

Chain BaB:  100%

BGC1  
GAL2  
SIA3

- Molecule 2: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-4)-beta-D-glucopyranose

Chain CaC:  33% 67%

BGC3  
GAL2  
SIA3

- Molecule 3: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose

Chain A:  50% 50%

GAL1  
SIA2

## 4 Data and refinement statistics

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	150.15Å 96.64Å 128.56Å 90.00° 110.49° 90.00°	Depositor
Resolution (Å)	44.80 – 1.55	Depositor
% Data completeness (in resolution range)	99.8 (44.80-1.55)	Depositor
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.00 (at 1.55Å)	Xtrriage
Refinement program	REFMAC 5.8.0258	Depositor
R, $R_{free}$	0.167 , 0.191	Depositor
Wilson B-factor (Å <sup>2</sup> )	19.3	Xtrriage
Anisotropy	0.031	Xtrriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
Total number of atoms	12254	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.68% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GAL, SIA, BGC, PEG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	AAA	0.66	0/2096	0.78	0/2848
1	BBB	0.64	0/2101	0.77	0/2856
1	CCC	0.65	0/2232	0.80	0/3032
1	DDD	0.64	0/2194	0.79	0/2984
1	EEE	0.65	0/2101	0.77	0/2854
All	All	0.65	0/10724	0.78	0/14574

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	AAA	2040	0	1970	6	0
1	BBB	2045	0	1973	9	0
1	CCC	2169	0	2088	12	0
1	DDD	2137	0	2058	11	0
1	EEE	2050	0	1981	9	0
2	BaB	43	0	37	0	0
2	CaC	32	0	26	0	0
3	A	21	0	17	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	AAA	7	0	10	0	0
4	EEE	7	0	10	0	0
5	AAA	308	0	0	1	0
5	BBB	320	0	0	2	0
5	CCC	365	0	0	4	0
5	DDD	383	0	0	1	0
5	EEE	327	0	0	1	0
All	All	12254	0	10170	40	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 2.

The worst 5 of 40 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:CCC:184[B]:GLU:HG2	5:CCC:585:HOH:O	1.82	0.78
1:AAA:218:GLY:O	1:EEE:223:PRO:CB	2.33	0.76
1:BBB:73:ASN:ND2	5:BBB:301:HOH:O	2.23	0.72
1:AAA:82:VAL:O	1:AAA:192[B]:ASN:ND2	2.23	0.70
1:EEE:203:ASP:OD2	1:EEE:206[B]:ARG:HD3	1.96	0.65

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	AAA	260/272 (96%)	253 (97%)	7 (3%)	0	100   100
1	BBB	260/272 (96%)	252 (97%)	8 (3%)	0	100   100
1	CCC	277/272 (102%)	266 (96%)	11 (4%)	0	100   100
1	DDD	274/272 (101%)	266 (97%)	8 (3%)	0	100   100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	EEE	258/272 (95%)	250 (97%)	8 (3%)	0	100	100
All	All	1329/1360 (98%)	1287 (97%)	42 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	AAA	227/237 (96%)	226 (100%)	1 (0%)	91	82
1	BBB	229/237 (97%)	228 (100%)	1 (0%)	91	82
1	CCC	246/237 (104%)	244 (99%)	2 (1%)	81	66
1	DDD	240/237 (101%)	237 (99%)	3 (1%)	69	44
1	EEE	228/237 (96%)	227 (100%)	1 (0%)	91	82
All	All	1170/1185 (99%)	1162 (99%)	8 (1%)	86	69

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	EEE	220	GLU
1	DDD	123	ASN
1	DDD	94[A]	LEU
1	CCC	94	LEU
1	DDD	94[B]	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [i](#)

Of 8 monosaccharides modelled in this entry, 6 were used for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	SIA	A	2	3	20,20,21	0.68	0	24,28,31	1.02	2 (8%)
2	BGC	BaB	1	2	12,12,12	0.87	1 (8%)	17,17,17	1.02	1 (5%)
2	GAL	BaB	2	2	11,11,12	0.61	0	15,15,17	1.27	3 (20%)
2	SIA	BaB	3	2	20,20,21	0.96	1 (5%)	24,28,31	1.14	2 (8%)
2	GAL	CaC	2	2	11,11,12	0.72	0	15,15,17	2.39	5 (33%)
2	SIA	CaC	3	2	20,20,21	0.94	1 (5%)	24,28,31	1.31	2 (8%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	SIA	A	2	3	-	0/18/34/38	0/1/1/1
2	BGC	BaB	1	2	-	0/2/22/22	0/1/1/1
2	GAL	BaB	2	2	-	0/2/19/22	0/1/1/1
2	SIA	BaB	3	2	-	0/18/34/38	0/1/1/1
2	GAL	CaC	2	2	-	0/2/19/22	0/1/1/1
2	SIA	CaC	3	2	-	0/18/34/38	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	CaC	3	SIA	C3-C2	2.33	1.56	1.52
2	BaB	1	BGC	O1-C1	2.19	1.46	1.39

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	BaB	3	SIA	C3-C2	2.04	1.55	1.52

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	CaC	2	GAL	O3-C3-C2	5.88	121.25	109.99
2	CaC	2	GAL	C2-C3-C4	3.62	117.16	110.89
2	CaC	2	GAL	C3-C4-C5	3.59	116.64	110.24
2	CaC	3	SIA	O6-C2-C3	-3.28	105.95	110.46
2	CaC	2	GAL	C1-O5-C5	3.23	116.56	112.19

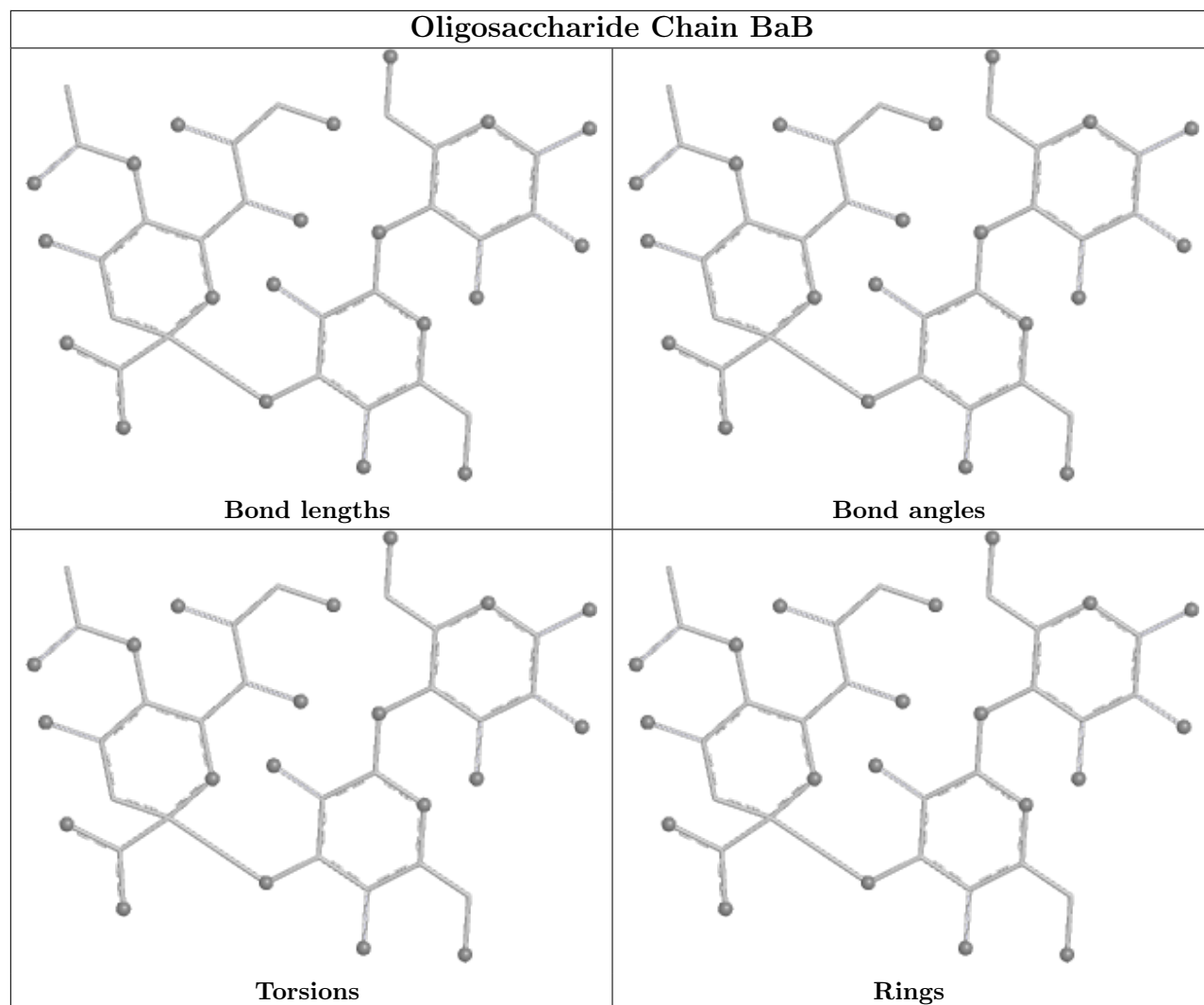
There are no chirality outliers.

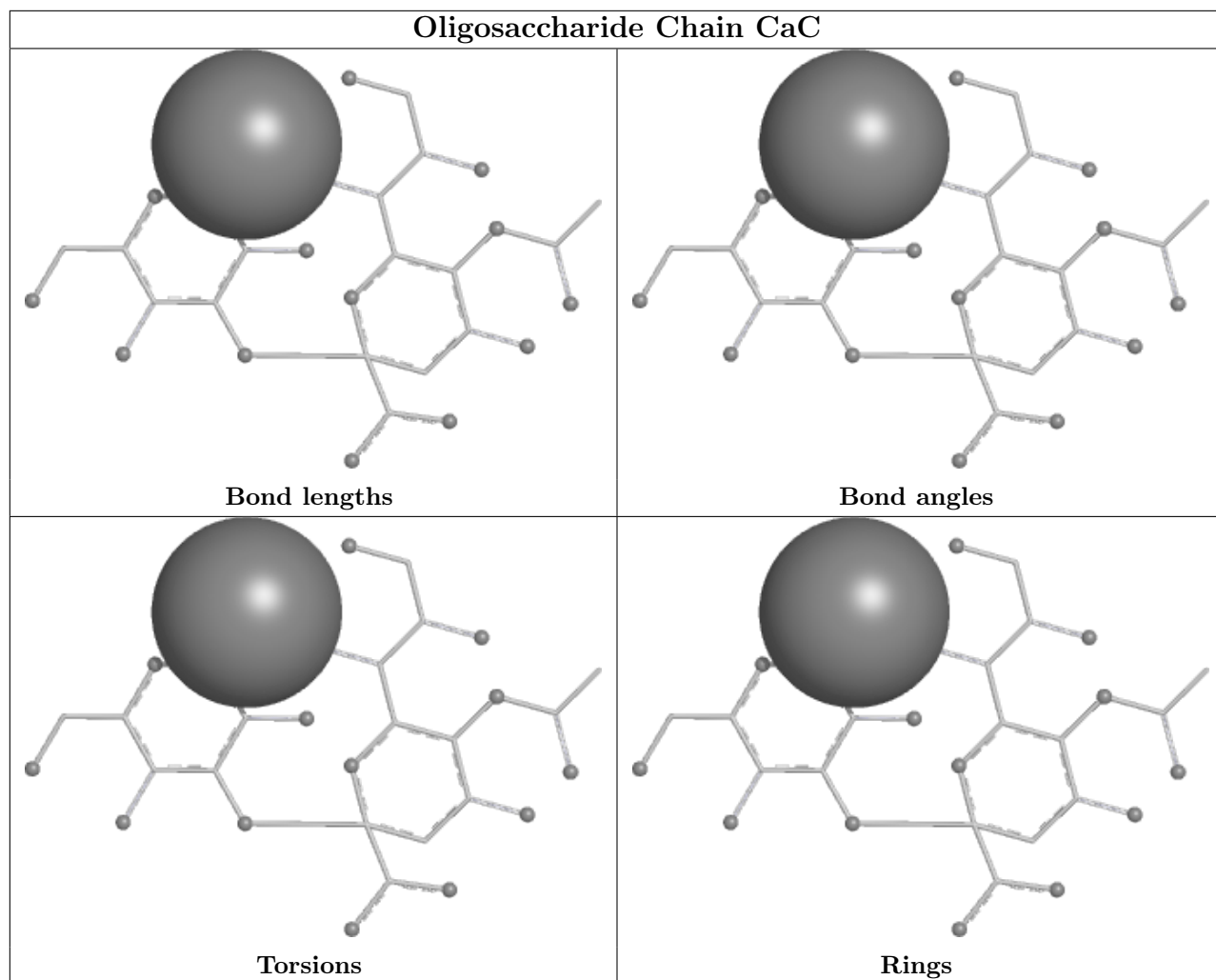
There are no torsion outliers.

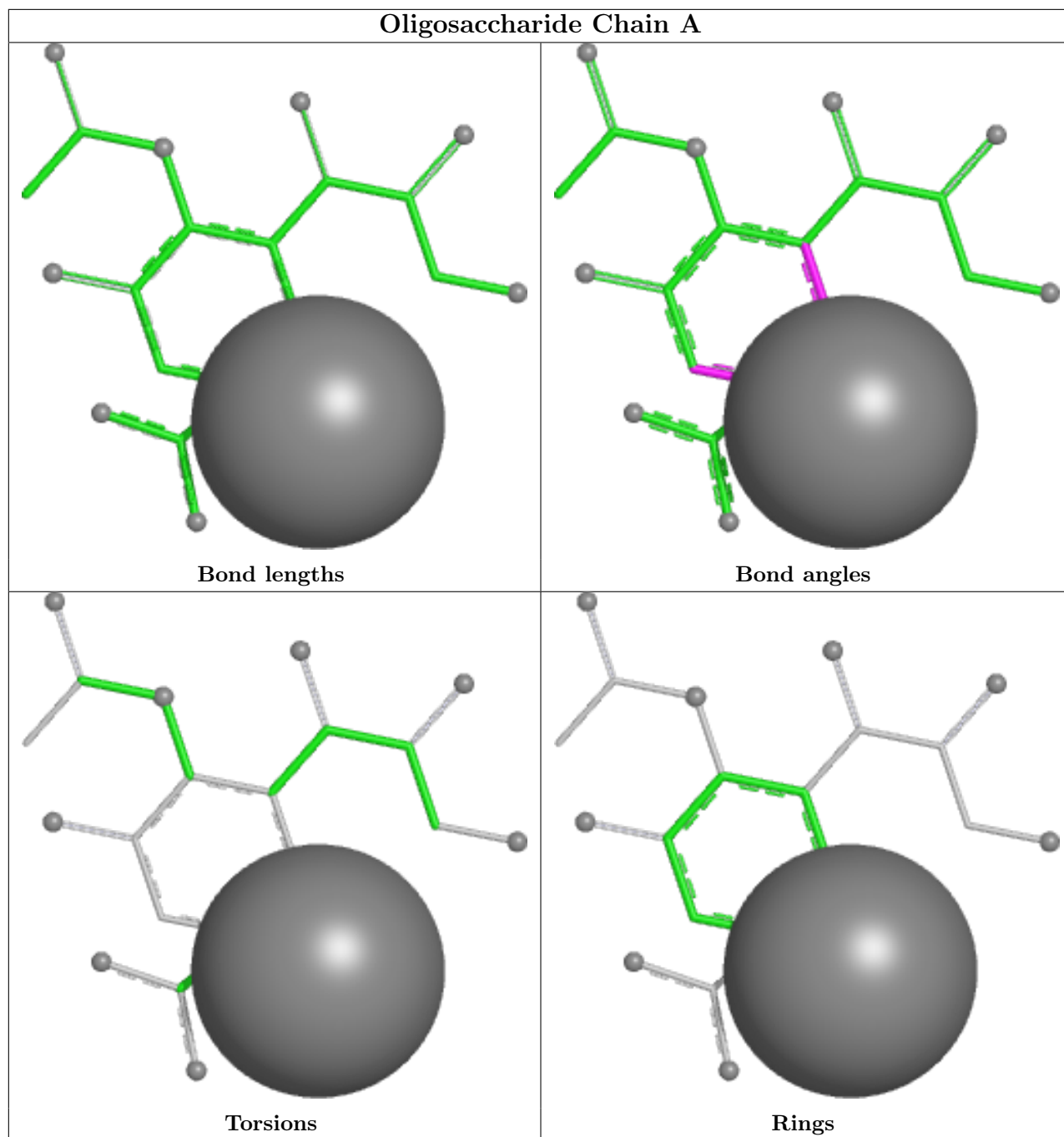
There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.







## 5.6 Ligand geometry [i](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection.



RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	PEG	AAA	601	-	6,6,6	0.19	0	5,5,5	0.14	0
4	PEG	EEE	501	-	6,6,6	0.22	0	5,5,5	0.14	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	PEG	AAA	601	-	-	3/4/4/4	-
4	PEG	EEE	501	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	AAA	601	PEG	O1-C1-C2-O2
4	EEE	501	PEG	O2-C3-C4-O4
4	EEE	501	PEG	O1-C1-C2-O2
4	AAA	601	PEG	O2-C3-C4-O4
4	AAA	601	PEG	C1-C2-O2-C3

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

EDS failed to run properly - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

EDS failed to run properly - this section is therefore empty.

### 6.3 Carbohydrates [i](#)

EDS failed to run properly - this section is therefore empty.

### 6.4 Ligands [i](#)

EDS failed to run properly - this section is therefore empty.

### 6.5 Other polymers [i](#)

EDS failed to run properly - this section is therefore empty.